General comments: This paper presents the application of the neural network to on-line radiative transfer calculation for accelerating the operational aerosol height retrievals from TROPOMI measurements. I agree with other two reviewers; the results are very significant and interesting about the speed up of three orders of magnitude without insignificant change of the retrieval accuracy. However, I think that several parts of this paper should be revised before publication.

Specific comments:

Page 2, line 12-15: With TROPOMI ~ 50,000 pixels per orbit in many cases.

- I am not clear for this sentence, does it mean that ~ 6 % of TROPOOMI pixels are typically identified as aerosol contaminated pixels based on UV aerosol index (> 0) for retrieving aerosol heights ?
- I would like to suggest "operational retrievals are time restricted~ for retrieving aerosol layer height" to be revised like "The operational computation capability is much restricted for TROPOOMI recording approximately 1.4 million pixels within a single orbit where 50, 000 pixels are typically identify as aerosol contaminated pixels for retrieving aerosol layer height."

Page 2 Line 17: while ➔ whereas

Page 3 Line 23: scaled by ➔ constrained with

Page 3 Line 26-28: This cost function is also constrained with a priori knowledge of the state vector x. The final retrieval product of zaer and τ

Page 4 line 1: "The forward model is employed to simulate the measured reflectance spectrum with model parameter x as following" or delete "modeled"

Page 4 line 4: delete ")" from wavelength λ)

Where I and Eo represent the Earth radiance and solar irradiance, respectively, with the cosine of the solar zenith angle (θ_0) u_o

Page 4 line 5: employed to update the state vector as following

Page 4 line 7: where ~~i is the current iteration and~~ Ki is the matrix of derivatives (Jacobian) of the reflectance with respect to state vector parameters at the current iteration i.

Page 4 line 9: An iterative estimate is convergent to a solution if the relative changes in the state vector is less than ~.

Page 4 Line 10-11: The retrieval is decided to be failed if ~ /their respective boundary conditions by OE ➔ the respective boundary conditions

Page 4 Line 15: The forward model iteratively simulates TOA radiance spectra until the convergence of x^2 (Equation 1).

Page 4 Line 16: To define TOA reflectance, convolved high resolution reference solar spectrum onto the instrument's slit function is used instead of measured solar irradiance?

Page 5 Line 4 : While ➔ In spite that

Page 5 Line 5: (Sanders and de Haan, 2016) ➔ Sanders and de Haan, 2016

or preliminary experiments have ~~ (Sanders and de Haan, 2016).

Page 5 Line 5-6: the impact of ignoring RRS on the retrieval of aerosol layer height using the oxygen A-band are much smaller than that of other retrieval errors such as ??.

Page 5 Line 6-7: Due to this~ cross sections ➔ Therefore, RRS has been historically not simulated in the forward model of the KNMI aerosol layer height retrieval algorithm.

Page 5 Line 8-9: The retrieval of Zaer in the presence of clouds is still challenging (reference) and thereby is performed only for cloud-free cases masked when cloud fraction is less than 0.2. Compared to totally cloud-free scene, the retrieval errors of Zaer are considerably problematic when the measured scene is masked as clear-sky in the presence of optically thin cirrus (reference).

Page 5 Line 10 : "Instead, TROPOMI incorporates information from the VIIRS instrument to detect the

Presence of clouds in the measured scene, which are further on mentioned in the output product flags

Instead"

- This author did not describe how to identify cloudy scenes up to here in previous algorithm or instrument instead current TROPOMI algorithm.

- "TROPOMI" ➔ TROPOMI Zaer algorithm or TROPOMI cloud algorithm?

- "which are further on mentioned in the output product flags" ➔ What is output product here? Aerosol height or cloud product? Anyway output product flag contains information on how to identify a pixel as cloud using cloud product between TROPOOMI or VIIRS? Please specify more.

Page 5 Line 11: Please provide details on how to decide the fraction of the pixel containing aerosols.

Page 5 Line 15-16 : The aerosol scattering phase function ~ significantly more computations ➔ A Henyey-Greenstein model (Henyey and Greenstein, 1941) is used to parametrize the aerosol scattering phase function, which is one of the widely used approximations.

Page 5 Line 17: what is the fixed aerosol optical properties taken from AERONET data?

I am not clear about "the consequences of fixing them" based on the following sentence "GOME-2 spectra to show that the algorithm is robust against these model assumptions"

Page 5 Line 26: Actually, does NN-based algorithm include any complexity of the model thanks to the speed up? If not, please revise "In constrat ~ endeavor" to "The speed up of forward model simulation encourages increasing the complexity of simulation assumption ~~" and move to section 5.

Page 6 Line 6: modeled measured reflectance ➔ modeled reflectance

Page 6 Line 8: are derived from ➔ are taken from|, which provide ➔, including.

What is about meteorological input at surface level?

Page 6 Line 9: The various ~ parameters == > The various geophysical parameters

Page 6 Line 11: requires ➔ takes?

Page 7 Line 4: Kingma and Ba (2014)) ➔ Kingma and Ba (2014)

Page 9 Line 34: Because of this == > Therefore

Page 12 : (a) I am not clear why "augmented" is used as an adjective for the neural network. It looks better just to indicate it as "the neural network". (b) in this section, most of Figures are not directly

introduced such as ~~~(Figure 9a), ~~~~(Table 4). Please try to introduce a Figure directly and then give relative analysis.

Page 12 Line 2: Figure 8a ➔ Figure 8b | insert Figure a before ", which" at line 1. Please use Figure 8b more in this analysis.

Page 12 line 2-3: absorbing aerosol index (AAI)

Page 12 line 4: Pixels that were cloud contaminated ➔ cloud-contaminated pixels

Page 12 line 4: What is the processing chain?

Page 12 line 5-6:

-Scientific comment: the FRESCO-based cloud fraction is positively biased for all typed aerosols or just for biomass burning aerosols. Please clarify.

-Editing comment: "However, the cloud-free biomass burning aerosol pixels could be screened out as the high cloud fraction of greater than the threshold is likely to be retrieved."

Page 12 line 7: ", as the surface ~ in these regions" ➔ where the surface albedo retrieval is likely to be wrong.

Page 12, line 3-7: This part should be described in detail in section 2: the TROPOMI aerosol layer height retrieval algorithm.

Page 12, line 8-10: Figure 9 compares the retrieved Zaer over the plume using the line-by-line and neural network based forward models, respectively. The number of the converged retrievals is 7418 for the line-by-line algorithm, but 7370 for the neural network algorithm.

Page 12, line 11: this analysis is contradictory, please revise and give more interprets on Fig 8.c; for example, where/why the positive/negative biases are dominant.

Page 12, line13-15: please revise this sentence, it is very hard to see what is the subject for "indicate" after respectively.

Page 12, line 18: due to over-estimation ➔ caused by over-estimation of ?? by.

Page 12, line 19: a consistent bias of 60 meters with a standard deviation of 30 meters.

Page 13, line 2: the aerosol layer height algorithm among L2 algorithms is unique for implementing on-line RT?

Page 13, line 4: Disamar just calculate radiance?

Page 13, line 22: We evaluate the Zaer retrieved from TROPOMI measurements over Southern California on 12 December 2017 when the fire plume extensively floats from land to ocean over a dry and almost cloudless scene.

Table 4 caption: Statistics of difference in retrieved zaer between Disamar and NN from figure 9c.

Figure 2, Figure 3, Figure 4: characters looks vague.